

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

REMARKS

Claims 1-12 and 14-23 are pending in the application.

Claims 1-12 and 14-23 have been rejected.

Reconsideration and allowance of the Claims is respectfully requested.

I. REJECTIONS UNDER 35 U.S.C. § 103

Claims 1, 4-6 and 9-12 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Sharman (US 5,744,854) in view of Hata, et al. (US 5,878,393) and "New Riverside University Dictionary" ("DIC").

Claims 2-3 and 21-23 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Sharman (US 5,744,854) in view of Hata, et al. (US 5,878,393) and DIC, and further in view of Oh (US 6,141,642).

Claim 7 was rejected under 35 U.S.C. § 103(a) as being unpatentable over Sharman (US 5,744,854) in view of Hata, et al. (US 5,878,393) and DIC, and further in view of Microsoft Press, "Computer Dictionary", page 298 ("R1").

Claim 8 was rejected under 35 U.S.C. § 103(a) as being unpatentable over Sharman (US 5,744,854) in view of Hata, et al. (US 5,878,393) and DIC, and R1, and further in view of O'Donnell ("Programming For The World - A Guide To Internationalization", ISBN 0-13-722190-8).).

Claims 14-15 and 19-20 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Sharman (US 5,744,854) in view of Hata, et al. (US 5,878,393) and DIC, and further in view of Malsheen, et al (US 4,979,216).

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

Claims 16 and 18 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Sharman (US 5,744,854) in view of Hata, et al. (US 5,878,393) and DIC and R1, and further in view of Oh (US 6,141,642).

Claim 17 was rejected under 35 U.S.C. § 103(a) as being unpatentable over Sharman (US 5,744,854) in view of Hata, et al. (US 5,878,393) and DIC, and R1, and O'Donnell ("Programming For The World - A Guide To Internationalization", ISBN 0-13-722190-8), and further in view of Oh (US 6,141,642).

The rejections are respectfully traversed.

Applicant again respectfully notes that to establish a *prima facie* case of obviousness, three basic criteria must be met. First, there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or to combine reference teachings. Second, there must be a reasonable expectation of success. Finally, the prior art reference (or references when combined) must teach or suggest all the claim limitations. Importantly, the teaching or suggestion to make the claimed invention and the reasonable expectation of success must both be found in the prior art, and not based on applicant's disclosure. MPEP § 2142.

Applicant respectfully submits, as detailed hereafter, that the Examiner's suggested combinations of references are improper, as they rely on selective hindsight reconstruction guided by Applicant's Claims and disclosure – not on the prospective teachings and suggestions of the cited references.

a. Claim 1

Applicant respectfully traverses the Examiner's suggested interpretations of the cited references.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

Applicant respectfully submits that Sharman discloses a text to speech system for converting input text into an output acoustic signal simulating natural speech. (Col. 1, lines 65-67). Sharman discloses breaking down words into constituent syllables. A dictionary may be used for the purpose of determining syllabic breaks. (Col. 5, lines 22-25). Constituent syllables are then further broken down into constituent phonemes. A dictionary look-up table – or some general purpose rules – may be used for this purpose. (Col. 5, lines 22-25). Sharman discloses removal and disregarding of any possible prefix or suffix, so as to enable the disaggregation of an underlying word into phonemes. (Col. 5, lines 26-29).

Sharman discloses annotation of phonemes with certain characteristics (e.g., pitch, duration). (Col. 5, lines 22-25). Steps are then performed to assemble phonemes into "breath groups." (Col. 5, line 48 – Col. 6, line 16). After the constituent phonemes and their characteristics have been determined, the acoustic processor determines diphones from the constituent phonemes. Each diphone represents a transition between two phonemes. A diphone library (with prerecorded sounds of the diphones) is accessed to retrieve the corresponding diphone samples, which are then concatenated to produce the output signal. See, generally, Col. 5, line 18-40; Col. 6, lines 22-38.

Therefore, Sharman parses the text file down to constituent phonemes. A group of phonemes are matched to diphones from a diphone library. Sound units are then grouped and generated using diphones. Sharman suggests that its approach is desirable to avoid excess processing. (Col. 2, lines 22-33).

The entire Sharman reference teaches away from processing whole words.

Thus far, the Examiner has conceded that: "Sharman fails to explicitly disclose utilizing 'speech sample' for the speech in the diphone library for the phonetic data"; "Sharman does not expressly disclose 'each speech sample corresponding to a one of said words, ... in said

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

vocabulary””; “Sharman does not expressly disclose ‘each speech sample corresponding to a one of said, ... prefixes and suffixes in said vocabulary’”; “Sharman does not expressly disclose ‘wherein said vocabulary of textural [sic] unit comprises words ... each having a pre-recorded speech sample associated therewith’”; and “Sharman does not expressly disclose ‘wherein said vocabulary of textural [sic] unit comprises ... prefixes and suffixes each having a pre-recorded speech sample associated therewith.’”

Applicant agrees with these concessions.

By comparison, Hata discloses a high-quality concatenative reading system for converting an input string into a sequence for subsequent audible synthesis. (Col. 1, lines 64-66). Hata discloses a dictionary of words and a word list generator coupled to the dictionary. The word list generator receives the input string and builds a word list from words stored in the dictionary corresponding to the input string. The word list generator assigns one or more prosodic environment tokens to word list entries – preferably to each entry in the word list. A reading system analyzes the word list to determine phonological features on the entries. Based on the word list, the environment token(s), and the phonological features, the reading system selects speech samples to be concatenated to supply the signal for audible synthesis. (Col. 1, line 66 – Col. 2, line 41).

Hata discloses a dictionary of digitally sampled sounds that have been recorded and stored in advance. (Col. 3, lines 42-44). The dictionary includes different samples for each possible pitch contour (i.e., prosodic environment) for each word in the dictionary. (Col. 4, lines 28-31). In addition to storing an entry for each prosodic environment of each word, the dictionary may also store all pronunciation variants of each word for each prosodic environment. (Col. 4, lines 37-55). Hata’s dictionary thus comprises multiple variations of each word entry.

Hata discloses input of text to be converted to speech. (Col. 4, lines 58-63). Hata discloses

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

a word list generator including a prosodic environment table that identifies the possible prosodic states of the text to speech implementation. (Col. 4, lines 63-65). The word list generator builds a word list comprising a word token, and a prosodic token for each word token, arranged in the order that they will be pronounced in the output speech. (Col. 5, lines 10-15). A reading system includes a phonological feature analyzer. For each word in the word list, the phonological feature analyzer evaluates the preceding and following words. Based on this information, the phonological feature analyzer selects a pre-recorded sample from the library having the corresponding prosodic and phonological characteristics. This is then added to a sample list, which is eventually output as the speech signal. (Col. 5, lines 16-31).

Regarding Claim 1, the Examiner postulates that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to combine Sharman and Hata to provide a stored speech sample in a word or other larger units (the removed prefix or suffix may be good candidate units, since they must associate some pronunciation unit for outputting, anyway)."

Applicant respectfully disagrees.

Again, the entire Sharman reference teaches away from processing whole words.

In order to combine the Sharman and Hata references as the Examiner has suggested – without the benefit of hindsight – one of ordinary skill in the art would have to: 1) find and read the Sharman reference; 2) understand Sharman's phoneme and diphone-based structures and operations; 3) spontaneously decide – in spite of Sharman's teaching away from processing whole words – that, out of all of Sharman's system, only its elemental and basic stored diphones needed to be replaced with stored speech samples in word or larger units; 4) seek out and find Hata's system; 5) disregard all of Hata's structures and methods, except for its library of pre-recorded samples for a given word and each tonal variation thereof; 6) *selectively* cull from Hata *only* its extensive library of pre-

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

recorded samples for a given word and each tonal variation thereof; 7) disregard Sharman's expressed intent to avoid excess processing; 8) replace Sharman's diphone library with Hata's extensive library of pre-recorded samples for a given word and each tonal variation thereof; 9) discard Sharman's extensive teaching of, and structure and operations for, breaking words down into their constituent syllables; 10) discard Sharman's structure and operations for phoneme duration assignment; 11) discard Sharman's structure and operations for breath group assembly; and 12) modify Sharman's remaining diphone-based structures and operations to successfully process and output Hata's pre-recorded word samples.

Applicant respectfully submits that there is no teaching or motivation in either Sharman or Hata to suggest such selective and substantial modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such selective and substantial modification.

The Examiner then goes on to speculate that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Sharman in view of Hata for specifically providing a mechanism to treat a prefix or suffix as same way [sic] as a word in a dictionary, as taught by DIC, so that each stored speech sample can correspond to one of word, prefix and suffix, for the purpose of selecting the appropriated [sic] 'granularity' or dictionary entry size to suit the specific application."

Applicant respectfully disagrees.

In order to further modify the already substantially modified Sharman/Hata combination as the Examiner has suggested – without the benefit of hindsight – one of ordinary skill in the art would have to: 1) spontaneously decide that – to the extent that either Sharman or Hata disclose or suggest treatment of prefixes or suffixes – both references were somehow deficient or incomplete with regard

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

thereto; 2) seek out and find the DIC reference; 3) arbitrarily decide, based upon a dictionary entry for either a prefix or suffix, that prefixes and suffixes should be treated as words; and 4) further modify the structures and operations of the already substantially modified Sharman/Hata combination to handle prefixes and suffixes as words.

Again, Applicant respectfully submits that there is no teaching or motivation in the cited references to suggest such highly selective and highly speculative modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such modification.

Furthermore, such selective and substantial modifications still fail to teach or suggest all the limitations of Claim 1.

Applicant respectfully submits that Claim 1 overcomes the rejection based upon a highly speculative and selective combination the Sharman and Hata references. Claim 1 is allowable. Applicant respectfully requests reconsideration and allowance of Claim 1.

b. Claims 4 and 5

Claims 4 and 5 depend from allowable Claim 1, and provide further limitations not taught or suggested by either Sharman or Hata.

Claim 4, and Claim 5 depending therefrom, requires creation of a speech unit by splicing together a plurality of speech samples prior to appending the speech unit to the output signal.

Claims 4 and 5 are thus allowable. Applicant respectfully requests reconsideration and allowance of Claims 4 and 5.

c. Claim 6

Applicant respectfully traverses the Examiner's suggested interpretations of the cited references.

Applicant again respectfully submits that Sharman discloses a text to speech system for

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

converting input text into an output acoustic signal simulating natural speech. (Col. 1, lines 65-67). Sharman discloses breaking down words into constituent syllables. A dictionary may be used for the purpose of determining syllabic breaks. (Col. 5, lines 22-25). Constituent syllables are then further broken down into constituent phonemes. A dictionary look-up table – or some general purpose rules – may be used for this purpose. (Col. 5, lines 22-25). Sharman discloses removal and disregarding of any possible prefix or suffix, so as to enable the disaggregation of an underlying word into phonemes. (Col. 5, lines 26-29).

Sharman discloses annotation of phonemes with certain characteristics (e.g., pitch, duration). (Col. 5, lines 22-25). Steps are then performed to assemble phonemes into “breath groups.” (Col. 5, line 48 – Col. 6, line 16). After the constituent phonemes and their characteristics have been determined, the acoustic processor determines diphones from the constituent phonemes. Each diphone represents a transition between two phonemes. A diphone library (with prerecorded sounds of the diphones) is accessed to retrieve the corresponding diphone samples, which are then concatenated to produce the output signal. See, generally, Col. 5, line 18-40; Col. 6, lines 22-38.

Therefore, Sharman parses the text file down to constituent phonemes. A group of phonemes are matched to diphones from a diphone library. Sound units are then grouped and generated using diphones. Sharman suggests that its approach is desirable to avoid excess processing. (Col. 2, lines 22-33).

The entire Sharman reference teaches away from processing whole words.

Thus far, the Examiner has conceded that: “Sharman fails to explicitly disclose utilizing ‘speech sample’ for the speech in the diphone library for the phonetic data”; “Sharman does not expressly disclose ‘each speech sample corresponding to a one of said words, ... in said vocabulary’”; “Sharman does not expressly disclose ‘each speech sample corresponding to a one of

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

said, ... prefixes and suffixes in said vocabulary''; ''Sharman does not expressly disclose 'wherein said vocabulary of textural [sic] unit comprises words ... each having a pre-recorded speech sample associated therewith''; and ''Sharman does not expressly disclose 'wherein said vocabulary of textural [sic] unit comprises ... prefixes and suffixes each having a pre-recorded speech sample associated therewith.'''

Applicant agrees with these concessions.

By comparison, Hata discloses a high-quality concatenative reading system for converting an input string into a sequence for subsequent audible synthesis. (Col. 1, lines 64-66). Hata discloses a dictionary of words and a word list generator coupled to the dictionary. The word list generator receives the input string and builds a word list from words stored in the dictionary corresponding to the input string. The word list generator assigns one or more prosodic environment tokens to word list entries – preferably to each entry in the word list. A reading system analyzes the word list to determine phonological features on the entries. Based on the word list, the environment token(s), and the phonological features, the reading system selects speech samples to be concatenated to supply the signal for audible synthesis. (Col. 1, line 66 – Col. 2, line 41).

Hata discloses a dictionary of digitally sampled sounds that have been recorded and stored in advance. (Col. 3, lines 42-44). The dictionary includes different samples for each possible pitch contour (i.e., prosodic environment) for each word in the dictionary. (Col. 4, lines 28-31). In addition to storing an entry for each prosodic environment of each word, the dictionary may also store all pronunciation variants of each word for each prosodic environment. (Col. 4, lines 37-55). Hata's dictionary thus comprises multiple variations of each word entry.

Hata discloses input of text to be converted to speech. (Col. 4, lines 58-63). Hata discloses a word list generator including a prosodic environment table that identifies the possible prosodic

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

states of the text to speech implementation. (Col. 4, lines 63-65). The word list generator builds a word list comprising a word token, and a prosodic token for each word token, arranged in the order that they will be pronounced in the output speech. (Col. 5, lines 10-15). A reading system includes a phonological feature analyzer. For each word in the word list, the phonological feature analyzer evaluates the preceding and following words. Based on this information, the phonological feature analyzer selects a pre-recorded sample from the library having the corresponding prosodic and phonological characteristics. This is then added to a sample list, which is eventually output as the speech signal. (Col. 5, lines 16-31).

Regarding Claim 6, the Examiner postulates that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Sharman for specifically providing a stored speech sample corresponding to a word or other unit, as taught by Hata, for the purpose of selecting the appropriated [sic] 'granularity' or dictionary entry size to suit the specific application."

Applicant respectfully disagrees.

Again, the entire Sharman reference teaches away from processing whole words.

In order to combine the Sharman and Hata references as the Examiner has suggested – without the benefit of hindsight – one of ordinary skill in the art would have to: 1) find and read the Sharman reference; 2) understand Sharman's phoneme and diphone-based structures and operations; 3) spontaneously decide – in spite of Sharman's teaching away from processing whole words – that, out of all of Sharman's system, only its elemental and basic stored diphones needed to be improved to include larger stored speech samples in word or larger units; 4) seek out and find Hata's system; 5) disregard all of Hata's structures and methods, except for its library of pre-recorded samples for a given word and each tonal variation thereof; 6) *selectively* cull from Hata *only* its extensive library

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

of pre-recorded samples for a given word and each tonal variation thereof; 7) disregard Sharman's expressed intent to avoid excess processing; 8) augment Sharman's diphone library with Hata's extensive library of pre-recorded samples for a given word and each tonal variation thereof; 9) selectively discard Sharman's extensive teaching of, and structure and operations for, breaking words down into their constituent syllables; 10) selectively discard Sharman's structure and operations for phoneme duration assignment; 11) selectively discard Sharman's structure and operations for breath group assembly; and 12) substantially modify and supplement Sharman's syllabic, phoneme and diphone-based structures and operations to successfully process and output Hata's pre-recorded word samples.

Applicant respectfully submits that there is no teaching or motivation in either Sharman or Hata to suggest such selective and substantial modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such selective and substantial modification.

The Examiner then goes on to speculate that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Sharman in view of Hata for specifically providing a mechanism to treat a prefix or suffix as same way [sic] as a word in a dictionary, as taught by DIC, so that each stored speech sample can correspond to one of word, prefix and suffix, for the purpose of selecting the appropriated [sic] 'granularity' or dictionary entry size to suit the specific application."

Applicant respectfully disagrees.

In order to further modify the already substantially modified Sharman/Hata combination as the Examiner has suggested – without the benefit of hindsight – one of ordinary skill in the art would have to: 1) spontaneously decide that – to the extent that either Sharman or Hata disclose or suggest

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

treatment of prefixes or suffixes – both references were somehow deficient or incomplete with regard thereto; 2) seek out and find the DIC reference; 3) arbitrarily decide, based upon a dictionary entry for either a prefix or suffix, that prefixes and suffixes should be treated as words; and 4) further modify the structures and operations of the already substantially modified Sharman/Hata combination to handle prefixes and suffixes as words.

Again, Applicant respectfully submits that there is no teaching or motivation in the cited references to suggest such highly selective and highly speculative modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such modification.

Furthermore, such selective and substantial modifications still fail to teach or suggest all the limitations of Claim 6.

Applicant respectfully submits that Claim 6 overcomes the rejection based upon a highly speculative and selective combination the Sharman, Hata and DIC references. Claim 6 is allowable. Applicant respectfully requests reconsideration and allowance of Claim 6.

d. Claim 9

Applicant respectfully traverses the Examiner's suggested interpretations of the cited references.

Applicant again respectfully submits that Sharman discloses a text to speech system for converting input text into an output acoustic signal simulating natural speech. (Col. 1, lines 65-67). Sharman discloses breaking down words into constituent syllables. A dictionary may be used for the purpose of determining syllabic breaks. (Col. 5, lines 22-25). Constituent syllables are then further broken down into constituent phonemes. A dictionary look-up table – or some general purpose rules – may be used for this purpose. (Col. 5, lines 22-25). Sharman discloses removal and disregarding of any possible prefix or suffix, so as to enable the disaggregation of an underlying

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

word into phonemes. (Col. 5, lines 26-29).

Sharman discloses annotation of phonemes with certain characteristics (e.g., pitch, duration). (Col. 5, lines 22-25). Steps are then performed to assemble phonemes into "breath groups." (Col. 5, line 48 – Col. 6, line 16). After the constituent phonemes and their characteristics have been determined, the acoustic processor determines diphones from the constituent phonemes. Each diphone represents a transition between two phonemes. A diphone library (with prerecorded sounds of the diphones) is accessed to retrieve the corresponding diphone samples, which are then concatenated to produce the output signal. See, generally, Col. 5, line 18-40; Col. 6, lines 22-38.

Therefore, Sharman parses the text file down to constituent phonemes. A group of phonemes are matched to diphones from a diphone library. Sound units are then grouped and generated using diphones. Sharman suggests that its approach is desirable to avoid excess processing. (Col. 2, lines 22-33).

The entire Sharman reference teaches away from processing whole words.

Thus far, the Examiner has conceded that: "Sharman fails to explicitly disclose utilizing 'speech sample' for the speech in the diphone library for the phonetic data"; "Sharman does not expressly disclose 'each speech sample corresponding to a one of said words, ... in said vocabulary'"; "Sharman does not expressly disclose 'each speech sample corresponding to a one of said, ... prefixes and suffixes in said vocabulary'"; "Sharman does not expressly disclose 'wherein said vocabulary of textural [sic] unit comprises words ... each having a pre-recorded speech sample associated therewith'"; and "Sharman does not expressly disclose 'wherein said vocabulary of textural [sic] unit comprises ... prefixes and suffixes each having a pre-recorded speech sample associated therewith.'"

Applicant agrees with these concessions.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

By comparison, Hata discloses a high-quality concatenative reading system for converting an input string into a sequence for subsequent audible synthesis. (Col. 1, lines 64-66). Hata discloses a dictionary of words and a word list generator coupled to the dictionary. The word list generator receives the input string and builds a word list from words stored in the dictionary corresponding to the input string. The word list generator assigns one or more prosodic environment tokens to word list entries – preferably to each entry in the word list. A reading system analyzes the word list to determine phonological features on the entries. Based on the word list, the environment token(s), and the phonological features, the reading system selects speech samples to be concatenated to supply the signal for audible synthesis. (Col. 1, line 66 – Col. 2, line 41).

Hata discloses a dictionary of digitally sampled sounds that have been recorded and stored in advance. (Col. 3, lines 42-44). The dictionary includes different samples for each possible pitch contour (i.e., prosodic environment) for each word in the dictionary. (Col. 4, lines 28-31). In addition to storing an entry for each prosodic environment of each word, the dictionary may also store all pronunciation variants of each word for each prosodic environment. (Col. 4, lines 37-55). Hata's dictionary thus comprises multiple variations of each word entry.

Hata discloses input of text to be converted to speech. (Col. 4, lines 58-63). Hata discloses a word list generator including a prosodic environment table that identifies the possible prosodic states of the text to speech implementation. (Col. 4, lines 63-65). The word list generator builds a word list comprising a word token, and a prosodic token for each word token, arranged in the order that they will be pronounced in the output speech. (Col. 5, lines 10-15). A reading system includes a phonological feature analyzer. For each word in the word list, the phonological feature analyzer evaluates the preceding and following words. Based on this information, the phonological feature analyzer selects a pre-recorded sample from the library having the corresponding prosodic and

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

phonological characteristics. This is then added to a sample list, which is eventually output as the speech signal. (Col. 5, lines 16-31).

Regarding Claim 9, the Examiner indicates that the rejection of the claim is based on the same reason as described for Claim 1.

Regarding Claim 1, the Examiner postulates that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to combine Sharman and Hata to provide a stored speech sample in a word or other larger units (the removed prefix or suffix may be good candidate units, since they must associate some pronunciation unit for outputting, anyway)."

Applicant respectfully disagrees.

Again, the entire Sharman reference teaches away from processing whole words.

In order to combine the Sharman and Hata references as the Examiner has suggested – without the benefit of hindsight – one of ordinary skill in the art would have to: 1) find and read the Sharman reference; 2) understand Sharman's phoneme and diphone-based structures and operations; 3) spontaneously decide – in spite of Sharman's teaching away from processing whole words – that, out of all of Sharman's system, only its elemental and basic stored diphones needed to be replaced with stored speech samples in word or larger units; 4) seek out and find Hata's system; 5) disregard all of Hata's structures and methods, except for its library of pre-recorded samples for a given word and each tonal variation thereof; 6) *selectively* cull from Hata *only* its extensive library of pre-recorded samples for a given word and each tonal variation thereof; 7) disregard Sharman's expressed intent to avoid excess processing; 8) replace Sharman's diphone library with Hata's extensive library of pre-recorded samples for a given word and each tonal variation thereof; 9) discard Sharman's extensive teaching of, and structure and operations for, breaking words down into their constituent syllables; 10) discard Sharman's structure and operations for phoneme duration

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

assignment; 11) discard Sharman's structure and operations for breath group assembly; and 12) modify Sharman's remaining diphone-based structures and operations to successfully process and output Hata's pre-recorded word samples.

Applicant respectfully submits that there is no teaching or motivation in either Sharman or Hata to suggest such selective and substantial modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such selective and substantial modification.

Furthermore, such selective and substantial modification still fails to teach or suggest all the limitations of Claim 1 and, subsequently, Claim 9.

Applicant respectfully submits that Claim 9 overcomes the rejection based upon a highly speculative and selective combination the Sharman and Hata references. Claim 9 is allowable. Applicant respectfully requests reconsideration and allowance of Claim 9.

e. Claim 10

Applicant respectfully traverses the Examiner's suggested interpretations of the cited references.

Applicant again respectfully submits that Sharman discloses a text to speech system for converting input text into an output acoustic signal simulating natural speech. (Col. 1, lines 65-67). Sharman discloses breaking down words into constituent syllables. A dictionary may be used for the purpose of determining syllabic breaks. (Col. 5, lines 22-25). Constituent syllables are then further broken down into constituent phonemes. A dictionary look-up table – or some general purpose rules – may be used for this purpose. (Col. 5, lines 22-25). Sharman discloses removal and disregarding of any possible prefix or suffix, so as to enable the disaggregation of an underlying word into phonemes. (Col. 5, lines 26-29).

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

Sharman discloses annotation of phonemes with certain characteristics (e.g., pitch, duration). (Col. 5, lines 22-25). Steps are then performed to assemble phonemes into "breath groups." (Col. 5, line 48 – Col. 6, line 16). After the constituent phonemes and their characteristics have been determined, the acoustic processor determines diphones from the constituent phonemes. Each diphone represents a transition between two phonemes. A diphone library (with prerecorded sounds of the diphones) is accessed to retrieve the corresponding diphone samples, which are then concatenated to produce the output signal. See, generally, Col. 5, line 18-40; Col. 6, lines 22-38.

Therefore, Sharman parses the text file down to constituent phonemes. A group of phonemes are matched to diphones from a diphone library. Sound units are then grouped and generated using diphones. Sharman suggests that its approach is desirable to avoid excess processing. (Col. 2, lines 22-33).

The entire Sharman reference teaches away from processing whole words.

Thus far, the Examiner has conceded that: "Sharman fails to explicitly disclose utilizing 'speech sample' for the speech in the diphone library for the phonetic data"; "Sharman does not expressly disclose 'each speech sample corresponding to a one of said words, ... in said vocabulary'"; "Sharman does not expressly disclose 'each speech sample corresponding to a one of said, ... prefixes and suffixes in said vocabulary'"; "Sharman does not expressly disclose 'wherein said vocabulary of textural [sic] unit comprises words ... each having a pre-recorded speech sample associated therewith'"; and "Sharman does not expressly disclose 'wherein said vocabulary of textural [sic] unit comprises ... prefixes and suffixes each having a pre-recorded speech sample associated therewith.'"

Applicant agrees with these concessions.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

By comparison, Hata discloses a high-quality concatenative reading system for converting an input string into a sequence for subsequent audible synthesis. (Col. 1, lines 64-66). Hata discloses a dictionary of words and a word list generator coupled to the dictionary. The word list generator receives the input string and builds a word list from words stored in the dictionary corresponding to the input string. The word list generator assigns one or more prosodic environment tokens to word list entries – preferably to each entry in the word list. A reading system analyzes the word list to determine phonological features on the entries. Based on the word list, the environment token(s), and the phonological features, the reading system selects speech samples to be concatenated to supply the signal for audible synthesis. (Col. 1, line 66 – Col. 2, line 41).

Hata discloses a dictionary of digitally sampled sounds that have been recorded and stored in advance. (Col. 3, lines 42-44). The dictionary includes different samples for each possible pitch contour (i.e., prosodic environment) for each word in the dictionary. (Col. 4, lines 28-31). In addition to storing an entry for each prosodic environment of each word, the dictionary may also store all pronunciation variants of each word for each prosodic environment. (Col. 4, lines 37-55). Hata's dictionary thus comprises multiple variations of each word entry.

Hata discloses input of text to be converted to speech. (Col. 4, lines 58-63). Hata discloses a word list generator including a prosodic environment table that identifies the possible prosodic states of the text to speech implementation. (Col. 4, lines 63-65). The word list generator builds a word list comprising a word token, and a prosodic token for each word token, arranged in the order that they will be pronounced in the output speech. (Col. 5, lines 10-15). A reading system includes a phonological feature analyzer. For each word in the word list, the phonological feature analyzer evaluates the preceding and following words. Based on this information, the phonological feature analyzer selects a pre-recorded sample from the library having the corresponding prosodic and

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

phonological characteristics. This is then added to a sample list, which is eventually output as the speech signal. (Col. 5, lines 16-31).

Regarding Claim 10, the Examiner indicates that the rejection of the claim is based on the same reason as described for Claim 1.

Regarding Claim 1, the Examiner postulates that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to combine Sharman and Hata to provide a stored speech sample in a word or other larger units (the removed prefix or suffix may be good candidate units, since they must associate some pronunciation unit for outputting, anyway)."

Applicant respectfully disagrees.

Again, the entire Sharman reference teaches away from processing whole words.

In order to combine the Sharman and Hata references as the Examiner has suggested – without the benefit of hindsight – one of ordinary skill in the art would have to: 1) find and read the Sharman reference; 2) understand Sharman's phoneme and diphone-based structures and operations; 3) spontaneously decide – in spite of Sharman's teaching away from processing whole words – that, out of all of Sharman's system, only its elemental and basic stored diphones needed to be replaced with stored speech samples in word or larger units; 4) seek out and find Hata's system; 5) disregard all of Hata's structures and methods, except for its library of pre-recorded samples for a given word and each tonal variation thereof; 6) *selectively* cull from Hata *only* its extensive library of pre-recorded samples for a given word and each tonal variation thereof; 7) disregard Sharman's expressed intent to avoid excess processing; 8) replace Sharman's diphone library with Hata's extensive library of pre-recorded samples for a given word and each tonal variation thereof; 9) discard Sharman's extensive teaching of, and structure and operations for, breaking words down into their constituent syllables; 10) discard Sharman's structure and operations for phoneme duration

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

assignment; 11) discard Sharman's structure and operations for breath group assembly; and 12) modify Sharman's remaining diphone-based structures and operations to successfully process and output Hata's pre-recorded word samples.

Applicant respectfully submits that there is no teaching or motivation in either Sharman or Hata to suggest such selective and substantial modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such selective and substantial modification.

Furthermore, such selective and substantial modification still fails to teach or suggest all the limitations of Claim 1 and, subsequently, Claim 10.

Applicant respectfully submits that Claim 10 overcomes the rejection based upon a highly speculative and selective combination the Sharman and Hata references. Claim 10 is allowable. Applicant respectfully requests reconsideration and allowance of Claim 10.

f. Claim 11

Applicant respectfully traverses the Examiner's suggested interpretations of the cited references.

Applicant again respectfully submits that Sharman discloses a text to speech system for converting input text into an output acoustic signal simulating natural speech. (Col. 1, lines 65-67). Sharman discloses breaking down words into constituent syllables. A dictionary may be used for the purpose of determining syllabic breaks. (Col. 5, lines 22-25). Constituent syllables are then further broken down into constituent phonemes. A dictionary look-up table – or some general purpose rules – may be used for this purpose. (Col. 5, lines 22-25). Sharman discloses removal and disregarding of any possible prefix or suffix, so as to enable the disaggregation of an underlying word into phonemes. (Col. 5, lines 26-29).

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

Sharman discloses annotation of phonemes with certain characteristics (e.g., pitch, duration). (Col. 5, lines 22-25). Steps are then performed to assemble phonemes into "breath groups." (Col. 5, line 48 – Col. 6, line 16). After the constituent phonemes and their characteristics have been determined, the acoustic processor determines diphones from the constituent phonemes. Each diphone represents a transition between two phonemes. A diphone library (with prerecorded sounds of the diphones) is accessed to retrieve the corresponding diphone samples, which are then concatenated to produce the output signal. See, generally, Col. 5, line 18-40; Col. 6, lines 22-38.

Therefore, Sharman parses the text file down to constituent phonemes. A group of phonemes are matched to diphones from a diphone library. Sound units are then grouped and generated using diphones. Sharman suggests that its approach is desirable to avoid excess processing. (Col. 2, lines 22-33).

The entire Sharman reference teaches away from processing whole words.

Thus far, the Examiner has conceded that: "Sharman fails to explicitly disclose utilizing 'speech sample' for the speech in the diphone library for the phonetic data"; "Sharman does not expressly disclose 'each speech sample corresponding to a one of said words, ... in said vocabulary'"; "Sharman does not expressly disclose 'each speech sample corresponding to a one of said, ... prefixes and suffixes in said vocabulary'"; "Sharman does not expressly disclose 'wherein said vocabulary of textural [sic] unit comprises words ... each having a pre-recorded speech sample associated therewith'"; and "Sharman does not expressly disclose 'wherein said vocabulary of textural [sic] unit comprises ... prefixes and suffixes each having a pre-recorded speech sample associated therewith.'"

Applicant agrees with these concessions.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

By comparison, Hata discloses a high-quality concatenative reading system for converting an input string into a sequence for subsequent audible synthesis. (Col. 1, lines 64-66). Hata discloses a dictionary of words and a word list generator coupled to the dictionary. The word list generator receives the input string and builds a word list from words stored in the dictionary corresponding to the input string. The word list generator assigns one or more prosodic environment tokens to word list entries – preferably to each entry in the word list. A reading system analyzes the word list to determine phonological features on the entries. Based on the word list, the environment token(s), and the phonological features, the reading system selects speech samples to be concatenated to supply the signal for audible synthesis. (Col. 1, line 66 – Col. 2, line 41).

Hata discloses a dictionary of digitally sampled sounds that have been recorded and stored in advance. (Col. 3, lines 42-44). The dictionary includes different samples for each possible pitch contour (i.e., prosodic environment) for each word in the dictionary. (Col. 4, lines 28-31). In addition to storing an entry for each prosodic environment of each word, the dictionary may also store all pronunciation variants of each word for each prosodic environment. (Col. 4, lines 37-55). Hata's dictionary thus comprises multiple variations of each word entry.

Hata discloses input of text to be converted to speech. (Col. 4, lines 58-63). Hata discloses a word list generator including a prosodic environment table that identifies the possible prosodic states of the text to speech implementation. (Col. 4, lines 63-65). The word list generator builds a word list comprising a word token, and a prosodic token for each word token, arranged in the order that they will be pronounced in the output speech. (Col. 5, lines 10-15). A reading system includes a phonological feature analyzer. For each word in the word list, the phonological feature analyzer evaluates the preceding and following words. Based on this information, the phonological feature analyzer selects a pre-recorded sample from the library having the corresponding prosodic and

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

phonological characteristics. This is then added to a sample list, which is eventually output as the speech signal. (Col. 5, lines 16-31).

Regarding Claim 11, the Examiner indicates that the rejection of the claim is based on the same reason as described for Claim 1.

Regarding Claim 1, the Examiner postulates that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to combine Sharman and Hata to provide a stored speech sample in a word or other larger units (the removed prefix or suffix may be good candidate units, since they must associate some pronunciation unit for outputting, anyway)."

Applicant respectfully disagrees.

Again, the entire Sharman reference teaches away from processing whole words.

In order to combine the Sharman and Hata references as the Examiner has suggested – without the benefit of hindsight – one of ordinary skill in the art would have to: 1) find and read the Sharman reference; 2) understand Sharman's phoneme and diphone-based structures and operations; 3) spontaneously decide – in spite of Sharman's teaching away from processing whole words – that, out of all of Sharman's system, only its elemental and basic stored diphones needed to be replaced with stored speech samples in word or larger units; 4) seek out and find Hata's system; 5) disregard all of Hata's structures and methods, except for its library of pre-recorded samples for a given word and each tonal variation thereof; 6) *selectively* cull from Hata *only* its extensive library of pre-recorded samples for a given word and each tonal variation thereof; 7) disregard Sharman's expressed intent to avoid excess processing; 8) replace Sharman's diphone library with Hata's extensive library of pre-recorded samples for a given word and each tonal variation thereof; 9) discard Sharman's extensive teaching of, and structure and operations for, breaking words down into their constituent syllables; 10) discard Sharman's structure and operations for phoneme duration

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

assignment; 11) discard Sharman's structure and operations for breath group assembly; and 12) modify Sharman's remaining diphone-based structures and operations to successfully process and output Hata's pre-recorded word samples.

Applicant respectfully submits that there is no teaching or motivation in either Sharman or Hata to suggest such selective and substantial modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such selective and substantial modification.

Furthermore, such selective and substantial modification still fails to teach or suggest all the limitations of Claim 1 and, subsequently, Claim 11.

Applicant respectfully submits that Claim 11 overcomes the rejection based upon a highly speculative and selective combination the Sharman and Hata references. Claim 11 is allowable. Applicant respectfully requests reconsideration and allowance of Claim 11.

g. Claim 12

Applicant respectfully traverses the Examiner's suggested interpretations of the cited references.

Applicant again respectfully submits that Sharman discloses a text to speech system for converting input text into an output acoustic signal simulating natural speech. (Col. 1, lines 65-67). Sharman discloses breaking down words into constituent syllables. A dictionary may be used for the purpose of determining syllabic breaks. (Col. 5, lines 22-25). Constituent syllables are then further broken down into constituent phonemes. A dictionary look-up table – or some general purpose rules – may be used for this purpose. (Col. 5, lines 22-25). Sharman discloses removal and disregarding of any possible prefix or suffix, so as to enable the disaggregation of an underlying word into phonemes. (Col. 5, lines 26-29).

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

Sharman discloses annotation of phonemes with certain characteristics (e.g., pitch, duration). (Col. 5, lines 22-25). Steps are then performed to assemble phonemes into "breath groups." (Col. 5, line 48 – Col. 6, line 16). After the constituent phonemes and their characteristics have been determined, the acoustic processor determines diphones from the constituent phonemes. Each diphone represents a transition between two phonemes. A diphone library (with prerecorded sounds of the diphones) is accessed to retrieve the corresponding diphone samples, which are then concatenated to produce the output signal. See, generally, Col. 5, line 18-40; Col. 6, lines 22-38.

Therefore, Sharman parses the text file down to constituent phonemes. A group of phonemes are matched to diphones from a diphone library. Sound units are then grouped and generated using diphones. Sharman suggests that its approach is desirable to avoid excess processing. (Col. 2, lines 22-33).

The entire Sharman reference teaches away from processing whole words.

Thus far, the Examiner has conceded that: "Sharman fails to explicitly disclose utilizing 'speech sample' for the speech in the diphone library for the phonetic data"; "Sharman does not expressly disclose 'each speech sample corresponding to a one of said words, ... in said vocabulary'"; "Sharman does not expressly disclose 'each speech sample corresponding to a one of said, ... prefixes and suffixes in said vocabulary'"; "Sharman does not expressly disclose 'wherein said vocabulary of textural [sic] unit comprises words ... each having a pre-recorded speech sample associated therewith'"; and "Sharman does not expressly disclose 'wherein said vocabulary of textural [sic] unit comprises ... prefixes and suffixes each having a pre-recorded speech sample associated therewith.'"

Applicant agrees with these concessions.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

By comparison, Hata discloses a high-quality concatenative reading system for converting an input string into a sequence for subsequent audible synthesis. (Col. 1, lines 64-66). Hata discloses a dictionary of words and a word list generator coupled to the dictionary. The word list generator receives the input string and builds a word list from words stored in the dictionary corresponding to the input string. The word list generator assigns one or more prosodic environment tokens to word list entries – preferably to each entry in the word list. A reading system analyzes the word list to determine phonological features on the entries. Based on the word list, the environment token(s), and the phonological features, the reading system selects speech samples to be concatenated to supply the signal for audible synthesis. (Col. 1, line 66 – Col. 2, line 41).

Hata discloses a dictionary of digitally sampled sounds that have been recorded and stored in advance. (Col. 3, lines 42-44). The dictionary includes different samples for each possible pitch contour (i.e., prosodic environment) for each word in the dictionary. (Col. 4, lines 28-31). In addition to storing an entry for each prosodic environment of each word, the dictionary may also store all pronunciation variants of each word for each prosodic environment. (Col. 4, lines 37-55). Hata's dictionary thus comprises multiple variations of each word entry.

Hata discloses input of text to be converted to speech. (Col. 4, lines 58-63). Hata discloses a word list generator including a prosodic environment table that identifies the possible prosodic states of the text to speech implementation. (Col. 4, lines 63-65). The word list generator builds a word list comprising a word token, and a prosodic token for each word token, arranged in the order that they will be pronounced in the output speech. (Col. 5, lines 10-15). A reading system includes a phonological feature analyzer. For each word in the word list, the phonological feature analyzer evaluates the preceding and following words. Based on this information, the phonological feature analyzer selects a pre-recorded sample from the library having the corresponding prosodic and

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

phonological characteristics. This is then added to a sample list, which is eventually output as the speech signal. (Col. 5, lines 16-31).

Regarding Claim 12, the Examiner indicates that the rejection of the claim is based on the same reason as described for Claims 1 and 6.

Regarding Claims 1 and 6, the Examiner postulates that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to combine Sharman and Hata to provide a stored speech sample in a word or other larger units (the removed prefix or suffix may be good candidate units, since they must associate some pronunciation unit for outputting, anyway)."

Applicant respectfully disagrees.

Again, the entire Sharman reference teaches away from processing whole words.

In order to combine the Sharman and Hata references as the Examiner has suggested – without the benefit of hindsight – one of ordinary skill in the art would have to: 1) find and read the Sharman reference; 2) understand Sharman's phoneme and diphone-based structures and operations; 3) spontaneously decide – in spite of Sharman's teaching away from processing whole words – that, out of all of Sharman's system, only its elemental and basic stored diphones needed to be improved to include larger stored speech samples in word or larger units; 4) seek out and find Hata's system; 5) disregard all of Hata's structures and methods, except for its library of pre-recorded samples for a given word and each tonal variation thereof; 6) *selectively* cull from Hata *only* its extensive library of pre-recorded samples for a given word and each tonal variation thereof; 7) disregard Sharman's expressed intent to avoid excess processing; 8) augment Sharman's diphone library with Hata's extensive library of pre-recorded samples for a given word and each tonal variation thereof; 9) selectively discard Sharman's extensive teaching of, and structure and operations for, breaking words down into their constituent syllables; 10) selectively discard Sharman's structure and operations for

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

phoneme duration assignment; 11) selectively discard Sharman's structure and operations for breath group assembly; and 12) substantially modify and supplement Sharman's syllabic, phoneme and diphone-based structures and operations to successfully process and output Hata's pre-recorded word samples.

Applicant respectfully submits that there is no teaching or motivation in either Sharman or Hata to suggest such selective and substantial modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such selective and substantial modification.

The Examiner then goes on to speculate that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Sharman in view of Hata for specifically providing a mechanism to treat a prefix or suffix as same way [sic] as a word in a dictionary, as taught by DIC, so that each stored speech sample can correspond to one of word, prefix and suffix, for the purpose of selecting the appropriated [sic] 'granularity' or dictionary entry size to suit the specific application."

Applicant respectfully disagrees.

In order to further modify the already substantially modified Sharman/Hata combination as the Examiner has suggested – without the benefit of hindsight – one of ordinary skill in the art would have to: 1) spontaneously decide that – to the extent that either Sharman or Hata disclose or suggest treatment of prefixes or suffixes – both references were somehow deficient or incomplete with regard thereto; 2) seek out and find the DIC reference; 3) arbitrarily decide, based upon a dictionary entry for either a prefix or suffix, that prefixes and suffixes should be treated as words; and 4) further modify the structures and operations of the already substantially modified Sharman/Hata combination to handle prefixes and suffixes as words.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

Again, Applicant respectfully submits that there is no teaching or motivation in the cited references to suggest such highly selective and highly speculative modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such modification.

Furthermore, such selective and substantial modifications still fail to teach or suggest all the limitations of Claims 1 and 6 and, subsequently, claim 12.

Applicant respectfully submits that Claim 12 overcomes the rejection based upon a highly speculative and selective combination the Sharman, Hata and DIC references. Claim 12 is allowable. Applicant respectfully requests reconsideration and allowance of Claim 12.

h. Claims 2 and 3

Claims 2 and 3 depend from allowable Claim 1, and provide further limitations not taught or suggested by either Sharman or Hata.

Claim 2, and Claim 3 depending therefrom, requires passing of an indicated textual unit to a secondary text to speech engine, and receiving a speech sample converted from the indicated textual unit from the secondary text to speech engine.

The Examiner has conceded that "Sharman in view of Hata and DIC does not expressly disclose 'passing said indicated textual unit to a secondary text to speech engine; receiving a speech sample converted from said indicated textual unit from said secondary text to speech engine'."

Applicant agrees.

The Examiner nonetheless goes on to suggest that one of ordinary skill in the art would be taught or motivated by a combination that does not disclose a secondary text to speech engine to: 1) seek out a fourth reference, the Oh reference; 2) disregard Oh extensive teaching of text to speech processing on a character-by-character basis; 3) selectively cull from Oh only a secondary text to speech processor; and 4) further modify the structure and operation of the already substantially

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

modified Sharman/Hata/DIC combination to integrate the secondary text to speech processor from Oh.

Applicant respectfully submits that there is no teaching or motivation in the cited references to suggest such highly selective and highly speculative modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such modification.

Furthermore, such selective and substantial modifications still fail to teach or suggest all the limitations of Claims 2 and 3.

Claims 2 and 3 are thus allowable. Applicant respectfully requests reconsideration and allowance of Claims 2 and 3.

i. Claim 21

Applicant respectfully traverses the Examiner's suggested interpretations of the cited references.

Applicant again respectfully submits that Sharman discloses a text to speech system for converting input text into an output acoustic signal simulating natural speech. (Col. 1, lines 65-67). Sharman discloses breaking down words into constituent syllables. A dictionary may be used for the purpose of determining syllabic breaks. (Col. 5, lines 22-25). Constituent syllables are then further broken down into constituent phonemes. A dictionary look-up table – or some general purpose rules – may be used for this purpose. (Col. 5, lines 22-25). Sharman discloses removal and disregarding of any possible prefix or suffix, so as to enable the disaggregation of an underlying word into phonemes. (Col. 5, lines 26-29).

Sharman discloses annotation of phonemes with certain characteristics (e.g., pitch, duration). (Col. 5, lines 22-25). Steps are then performed to assemble phonemes into “breath groups.” (Col. 5, line 48 – Col. 6, line 16). After the constituent phonemes and their characteristics have been

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

determined, the acoustic processor determines diphones from the constituent phonemes. Each diphone represents a transition between two phonemes. A diphone library (with prerecorded sounds of the diphones) is accessed to retrieve the corresponding diphone samples, which are then concatenated to produce the output signal. See, generally, Col. 5, line 18-40; Col. 6, lines 22-38.

Therefore, Sharman parses the text file down to constituent phonemes. A group of phonemes are matched to diphones from a diphone library. Sound units are then grouped and generated using diphones. Sharman suggests that its approach is desirable to avoid excess processing. (Col. 2, lines 22-33).

The entire Sharman reference teaches away from processing whole words.

Thus far, the Examiner has conceded that: "Sharman fails to explicitly disclose utilizing 'speech sample' for the speech in the diphone library for the phonetic data"; "Sharman does not expressly disclose 'each speech sample corresponding to a one of said words, ... in said vocabulary'"; "Sharman does not expressly disclose 'each speech sample corresponding to a one of said, ... prefixes and suffixes in said vocabulary'"; "Sharman does not expressly disclose 'wherein said vocabulary of textural [sic] unit comprises words ... each having a pre-recorded speech sample associated therewith'"; and "Sharman does not expressly disclose 'wherein said vocabulary of textural [sic] unit comprises ... prefixes and suffixes each having a pre-recorded speech sample associated therewith.'"

Applicant agrees with these concessions.

By comparison, Hata discloses a high-quality concatenative reading system for converting an input string into a sequence for subsequent audible synthesis. (Col. 1, lines 64-66). Hata discloses a dictionary of words and a word list generator coupled to the dictionary. The word list generator receives the input string and builds a word list from words stored in the dictionary

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

corresponding to the input string. The word list generator assigns one or more prosodic environment tokens to word list entries – preferably to each entry in the word list. A reading system analyzes the word list to determine phonological features on the entries. Based on the word list, the environment token(s), and the phonological features, the reading system selects speech samples to be concatenated to supply the signal for audible synthesis. (Col. 1, line 66 – Col. 2, line 41).

Hata discloses a dictionary of digitally sampled sounds that have been recorded and stored in advance. (Col. 3, lines 42-44). The dictionary includes different samples for each possible pitch contour (i.e., prosodic environment) for each word in the dictionary. (Col. 4, lines 28-31). In addition to storing an entry for each prosodic environment of each word, the dictionary may also store all pronunciation variants of each word for each prosodic environment. (Col. 4, lines 37-55). Hata's dictionary thus comprises multiple variations of each word entry.

Hata discloses input of text to be converted to speech. (Col. 4, lines 58-63). Hata discloses a word list generator including a prosodic environment table that identifies the possible prosodic states of the text to speech implementation. (Col. 4, lines 63-65). The word list generator builds a word list comprising a word token, and a prosodic token for each word token, arranged in the order that they will be pronounced in the output speech. (Col. 5, lines 10-15). A reading system includes a phonological feature analyzer. For each word in the word list, the phonological feature analyzer evaluates the preceding and following words. Based on this information, the phonological feature analyzer selects a pre-recorded sample from the library having the corresponding prosodic and phonological characteristics. This is then added to a sample list, which is eventually output as the speech signal. (Col. 5, lines 16-31).

Regarding Claim 21, the Examiner indicates that the rejection of the claim is based on the same reason as described for Claims 1, 2 and 6.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

Regarding Claims 1 and 6, the Examiner postulates that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to combine Sharman and Hata to provide a stored speech sample in a word or other larger units (the removed prefix or suffix may be good candidate units, since they must associate some pronunciation unit for outputting, anyway)."

Applicant respectfully disagrees.

Again, the entire Sharman reference teaches away from processing whole words.

In order to combine the Sharman and Hata references as the Examiner has suggested – without the benefit of hindsight – one of ordinary skill in the art would have to: 1) find and read the Sharman reference; 2) understand Sharman's phoneme and diphone-based structures and operations; 3) spontaneously decide – in spite of Sharman's teaching away from processing whole words – that, out of all of Sharman's system, only its elemental and basic stored diphones needed to be improved to include larger stored speech samples in word or larger units; 4) seek out and find Hata's system; 5) disregard all of Hata's structures and methods, except for its library of pre-recorded samples for a given word and each tonal variation thereof; 6) *selectively* cull from Hata *only* its extensive library of pre-recorded samples for a given word and each tonal variation thereof; 7) disregard Sharman's expressed intent to avoid excess processing; 8) augment Sharman's diphone library with Hata's extensive library of pre-recorded samples for a given word and each tonal variation thereof; 9) selectively discard Sharman's extensive teaching of, and structure and operations for, breaking words down into their constituent syllables; 10) selectively discard Sharman's structure and operations for phoneme duration assignment; 11) selectively discard Sharman's structure and operations for breath group assembly; and 12) substantially modify and supplement Sharman's syllabic, phoneme and diphone-based structures and operations to successfully process and output Hata's pre-recorded word samples.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

Applicant respectfully submits that there is no teaching or motivation in either Sharman or Hata to suggest such selective and substantial modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such selective and substantial modification.

The Examiner then goes on to speculate that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Sharman in view of Hata for specifically providing a mechanism to treat a prefix or suffix as same way [sic] as a word in a dictionary, as taught by DIC, so that each stored speech sample can correspond to one of word, prefix and suffix, for the purpose of selecting the appropriated [sic] 'granularity' or dictionary entry size to suit the specific application."

Applicant respectfully disagrees.

In order to further modify the already substantially modified Sharman/Hata combination as the Examiner has suggested – without the benefit of hindsight – one of ordinary skill in the art would have to: 1) spontaneously decide that – to the extent that either Sharman or Hata disclose or suggest treatment of prefixes or suffixes – both references were somehow deficient or incomplete with regard thereto; 2) seek out and find the DIC reference; 3) arbitrarily decide, based upon a dictionary entry for either a prefix or suffix, that prefixes and suffixes should be treated as words; and 4) further modify the structures and operations of the already substantially modified Sharman/Hata combination to handle prefixes and suffixes as words.

Again, Applicant respectfully submits that there is no teaching or motivation in the cited references to suggest such highly selective and highly speculative modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such modification.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

Claim 2 depends from allowable Claim 1, and provides further limitations not taught or suggested by either Sharman or Hata.

Claim 2 requires passing of an indicated textual unit to a secondary text to speech engine, and receiving a speech sample converted from the indicated textual unit from the secondary text to speech engine.

The Examiner has conceded that "Sharman in view of Hata and DIC does not expressly disclose 'passing said indicated textual unit to a secondary text to speech engine; receiving a speech sample converted from said indicated textual unit from said secondary text to speech engine'."

Applicant agrees.

The Examiner nonetheless goes on to suggest that one of ordinary skill in the art would be taught or motivated by a combination that does not disclose a secondary text to speech engine to: 1) seek out a fourth reference, the Oh reference; 2) disregard Oh extensive teaching of text to speech processing on a character-by-character basis; 3) selectively cull from Oh only a secondary text to speech processor; and 4) further modify the structure and operation of the already substantially modified Sharman/Hata/DIC combination to integrate the secondary text to speech processor from Oh.

Applicant respectfully submits that there is no teaching or motivation in the cited references to suggest such highly selective and highly speculative modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such modification.

Furthermore, such selective and substantial modifications still fail to teach or suggest all the limitations of Claims 1, 2, 6 and, subsequently, 21.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

Applicant respectfully submits that Claim 21 overcomes the rejection based upon a highly speculative and selective combination the Sharman, Hata, DIC and Oh references. Claim 21 is allowable. Applicant respectfully requests reconsideration and allowance of Claim 21.

j. Claim 22

Claim 22 depends from allowable Claim 21, and provides further limitations distinguishing over the cited references.

Claim 22 is thus allowable. Applicant respectfully requests reconsideration and allowance of Claim 22.

k. Claim 23

Applicant respectfully traverses the Examiner's suggested interpretations of the cited references.

Applicant again respectfully submits that Sharman discloses a text to speech system for converting input text into an output acoustic signal simulating natural speech. (Col. 1, lines 65-67). Sharman discloses breaking down words into constituent syllables. A dictionary may be used for the purpose of determining syllabic breaks. (Col. 5, lines 22-25). Constituent syllables are then further broken down into constituent phonemes. A dictionary look-up table – or some general purpose rules – may be used for this purpose. (Col. 5, lines 22-25). Sharman discloses removal and disregarding of any possible prefix or suffix, so as to enable the disaggregation of an underlying word into phonemes. (Col. 5, lines 26-29).

Sharman discloses annotation of phonemes with certain characteristics (e.g., pitch, duration). (Col. 5, lines 22-25). Steps are then performed to assemble phonemes into “breath groups.” (Col. 5, line 48 – Col. 6, line 16). After the constituent phonemes and their characteristics have been determined, the acoustic processor determines diphones from the constituent phonemes. Each

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

diphone represents a transition between two phonemes. A diphone library (with prerecorded sounds of the diphones) is accessed to retrieve the corresponding diphone samples, which are then concatenated to produce the output signal. See, generally, Col. 5, line 18-40; Col. 6, lines 22-38.

Therefore, Sharman parses the text file down to constituent phonemes. A group of phonemes are matched to diphones from a diphone library. Sound units are then grouped and generated using diphones. Sharman suggests that its approach is desirable to avoid excess processing. (Col. 2, lines 22-33).

The entire Sharman reference teaches away from processing whole words.

Thus far, the Examiner has conceded that: "Sharman fails to explicitly disclose utilizing 'speech sample' for the speech in the diphone library for the phonetic data"; "Sharman does not expressly disclose 'each speech sample corresponding to a one of said words, ... in said vocabulary'"; "Sharman does not expressly disclose 'each speech sample corresponding to a one of said, ... prefixes and suffixes in said vocabulary'"; "Sharman does not expressly disclose 'wherein said vocabulary of textural [sic] unit comprises words ... each having a pre-recorded speech sample associated therewith'"; and "Sharman does not expressly disclose 'wherein said vocabulary of textural [sic] unit comprises ... prefixes and suffixes each having a pre-recorded speech sample associated therewith.'"

Applicant agrees with these concessions.

By comparison, Hata discloses a high-quality concatenative reading system for converting an input string into a sequence for subsequent audible synthesis. (Col. 1, lines 64-66). Hata discloses a dictionary of words and a word list generator coupled to the dictionary. The word list generator receives the input string and builds a word list from words stored in the dictionary corresponding to the input string. The word list generator assigns one or more prosodic environment

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

tokens to word list entries – preferably to each entry in the word list. A reading system analyzes the word list to determine phonological features on the entries. Based on the word list, the environment token(s), and the phonological features, the reading system selects speech samples to be concatenated to supply the signal for audible synthesis. (Col. 1, line 66 – Col. 2, line 41).

Hata discloses a dictionary of digitally sampled sounds that have been recorded and stored in advance. (Col. 3, lines 42-44). The dictionary includes different samples for each possible pitch contour (i.e., prosodic environment) for each word in the dictionary. (Col. 4, lines 28-31). In addition to storing an entry for each prosodic environment of each word, the dictionary may also store all pronunciation variants of each word for each prosodic environment. (Col. 4, lines 37-55). Hata's dictionary thus comprises multiple variations of each word entry.

Hata discloses input of text to be converted to speech. (Col. 4, lines 58-63). Hata discloses a word list generator including a prosodic environment table that identifies the possible prosodic states of the text to speech implementation. (Col. 4, lines 63-65). The word list generator builds a word list comprising a word token, and a prosodic token for each word token, arranged in the order that they will be pronounced in the output speech. (Col. 5, lines 10-15). A reading system includes a phonological feature analyzer. For each word in the word list, the phonological feature analyzer evaluates the preceding and following words. Based on this information, the phonological feature analyzer selects a pre-recorded sample from the library having the corresponding prosodic and phonological characteristics. This is then added to a sample list, which is eventually output as the speech signal. (Col. 5, lines 16-31).

Regarding Claim 23, the Examiner indicates that the rejection of the claim is based on the same reason as described for Claims 1, 2 and 6.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

Regarding Claims 1 and 6, the Examiner postulates that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to combine Sharman and Hata to provide a stored speech sample in a word or other larger units (the removed prefix or suffix may be good candidate units, since they must associate some pronunciation unit for outputting, anyway)."

Applicant respectfully disagrees.

Again, the entire Sharman reference teaches away from processing whole words.

In order to combine the Sharman and Hata references as the Examiner has suggested – without the benefit of hindsight – one of ordinary skill in the art would have to: 1) find and read the Sharman reference; 2) understand Sharman's phoneme and diphone-based structures and operations; 3) spontaneously decide – in spite of Sharman's teaching away from processing whole words – that, out of all of Sharman's system, only its elemental and basic stored diphones needed to be improved to include larger stored speech samples in word or larger units; 4) seek out and find Hata's system; 5) disregard all of Hata's structures and methods, except for its library of pre-recorded samples for a given word and each tonal variation thereof; 6) *selectively* cull from Hata *only* its extensive library of pre-recorded samples for a given word and each tonal variation thereof; 7) disregard Sharman's expressed intent to avoid excess processing; 8) augment Sharman's diphone library with Hata's extensive library of pre-recorded samples for a given word and each tonal variation thereof; 9) selectively discard Sharman's extensive teaching of, and structure and operations for, breaking words down into their constituent syllables; 10) selectively discard Sharman's structure and operations for phoneme duration assignment; 11) selectively discard Sharman's structure and operations for breath group assembly; and 12) substantially modify and supplement Sharman's syllabic, phoneme and diphone-based structures and operations to successfully process and output Hata's pre-recorded word samples.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

Applicant respectfully submits that there is no teaching or motivation in either Sharman or Hata to suggest such selective and substantial modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such selective and substantial modification.

The Examiner then goes on to speculate that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Sharman in view of Hata for specifically providing a mechanism to treat a prefix or suffix as same way [sic] as a word in a dictionary, as taught by DIC, so that each stored speech sample can correspond to one of word, prefix and suffix, for the purpose of selecting the appropriated [sic] 'granularity' or dictionary entry size to suit the specific application."

Applicant respectfully disagrees.

In order to further modify the already substantially modified Sharman/Hata combination as the Examiner has suggested – without the benefit of hindsight – one of ordinary skill in the art would have to: 1) spontaneously decide that – to the extent that either Sharman or Hata disclose or suggest treatment of prefixes or suffixes – both references were somehow deficient or incomplete with regard thereto; 2) seek out and find the DIC reference; 3) arbitrarily decide, based upon a dictionary entry for either a prefix or suffix, that prefixes and suffixes should be treated as words; and 4) further modify the structures and operations of the already substantially modified Sharman/Hata combination to handle prefixes and suffixes as words.

Again, Applicant respectfully submits that there is no teaching or motivation in the cited references to suggest such highly selective and highly speculative modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such modification.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

Claim 2 depends from allowable Claim 1, and provides further limitations not taught or suggested by either Sharman or Hata.

Claim 2 requires passing of an indicated textual unit to a secondary text to speech engine, and receiving a speech sample converted from the indicated textual unit from the secondary text to speech engine.

The Examiner has conceded that "Sharman in view of Hata and DIC does not expressly disclose 'passing said indicated textual unit to a secondary text to speech engine; receiving a speech sample converted from said indicated textual unit from said secondary text to speech engine'."

Applicant agrees.

The Examiner nonetheless goes on to suggest that one of ordinary skill in the art would be taught or motivated by a combination that does not disclose a secondary text to speech engine to: 1) seek out a fourth reference, the Oh reference; 2) disregard Oh extensive teaching of text to speech processing on a character-by-character basis; 3) selectively cull from Oh only a secondary text to speech processor; and 4) further modify the structure and operation of the already substantially modified Sharman/Hata/DIC combination to integrate the secondary text to speech processor from Oh.

Applicant respectfully submits that there is no teaching or motivation in the cited references to suggest such highly selective and highly speculative modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such modification.

Furthermore, such selective and substantial modifications still fail to teach or suggest all the limitations of Claims 1, 2, 6 and, subsequently, 23.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

Applicant respectfully submits that Claim 23 overcomes the rejection based upon a highly speculative and selective combination the Sharman, Hata, DIC and Oh references. Claim 23 is allowable. Applicant respectfully requests reconsideration and allowance of Claim 23.

1. Claim 7

Claim 7 depends from allowable Claim 6, and provides further limitations not taught or suggested by Sharman, Hata or DIC.

Claim 7 requires *marking* of a parsed textual unit as being out of vocabulary, and adding the *marked textual unit* to the list.

The Examiner has previously conceded that "Sharman in view of Hata and DIC does not expressly disclose to mark a text unit that does not match the one either in dictionary or by rules sets."

Applicant agrees.

Nevertheless, the Examiner then contends that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Sharman by specifically marking a text unit of the processed data, as taught by R1, for the purpose of distinguishing the text unit that is not in the dictionary and preparing for further processing stages."

Despite this highly speculative assertion, Applicant finds no suggestion or motivation in Sharman, Hata or DIC for one of ordinary skill in the art to: 1) read the Sharman reference; 2) selectively and substantially modify Sharman by Hata; 3) selectively supplement the Sharman/Hata modification with a third reference, DIC; 4) spontaneously determine, after finding no teaching or suggestion of marking textual units in a substantial and highly selective combination of three references, that marking textual units is desirable or necessary; 5) find a fourth reference, R1; 6) selectively interpret the R1 reference to suggestion marking of textual units in the Sharman/Hata/DIC

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

combination; and 7) selectively modify the Sharman/Hata/DIC combination to add structure and operations for marking textual units.

Applicant respectfully submits that there is no teaching or motivation in either Sharman or Hata or DIC to suggest such selective and substantial modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such selective and substantial modification.

Furthermore, such selective and substantial modification still fails to teach or suggest all the limitations of Claim 7.

Applicant respectfully submits that Claim 7 overcomes the rejection based upon a highly speculative and selective combination of four references. Claim 7 is allowable. Applicant respectfully requests reconsideration and allowance of Claim 7.

m. Claim 8

Claim 8 depends from allowable Claim 7, and provides further limitations distinguishing over the cited references.

Applicant respectfully traverses the Examiner's interpretation of the cited references, as well as the selective and speculative combinations thereof.

Claim 8 is allowable. Applicant respectfully requests reconsideration and allowance of Claim 8.

n. Claim 14

Applicant respectfully traverses the Examiner's suggested interpretations of the cited references.

Applicant again respectfully submits that Sharman discloses a text to speech system for converting input text into an output acoustic signal simulating natural speech. (Col. 1, lines 65-67).

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

Sharman discloses breaking down words into constituent syllables. A dictionary may be used for the purpose of determining syllabic breaks. (Col. 5, lines 22-25). Constituent syllables are then further broken down into constituent phonemes. A dictionary look-up table – or some general purpose rules – may be used for this purpose. (Col. 5, lines 22-25). Sharman discloses removal and disregarding of any possible prefix or suffix, so as to enable the disaggregation of an underlying word into phonemes. (Col. 5, lines 26-29).

Sharman discloses annotation of phonemes with certain characteristics (e.g., pitch, duration). (Col. 5, lines 22-25). Steps are then performed to assemble phonemes into “breath groups.” (Col. 5, line 48 – Col. 6, line 16). After the constituent phonemes and their characteristics have been determined, the acoustic processor determines diphones from the constituent phonemes. Each diphone represents a transition between two phonemes. A diphone library (with prerecorded sounds of the diphones) is accessed to retrieve the corresponding diphone samples, which are then concatenated to produce the output signal. See, generally, Col. 5, line 18-40; Col. 6, lines 22-38.

Therefore, Sharman parses the text file down to constituent phonemes. A group of phonemes are matched to diphones from a diphone library. Sound units are then grouped and generated using diphones. Sharman suggests that its approach is desirable to avoid excess processing. (Col. 2, lines 22-33).

The entire Sharman reference teaches away from processing whole words.

Thus far, the Examiner has conceded that: “Sharman fails to explicitly disclose utilizing ‘speech sample’ for the speech in the diphone library for the phonetic data”; “Sharman does not expressly disclose ‘a data structure’ including several fields”; “Sharman does not expressly disclose ‘each speech sample corresponding to a one of said words, ... in said vocabulary’”; “Sharman does not expressly disclose ‘each speech sample corresponding to a one of said, ... prefixes and suffixes

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

in said vocabulary””; “Sharman does not expressly disclose ‘wherein said vocabulary of textual [sic] unit comprises words ... each having a pre-recorded speech sample associated therewith””; and “Sharman does not expressly disclose ‘wherein said vocabulary of textual [sic] unit comprises ... prefixes and suffixes each having a pre-recorded speech sample associated therewith.”

Applicant agrees with these concessions.

By comparison, Hata discloses a high-quality concatenative reading system for converting an input string into a sequence for subsequent audible synthesis. (Col. 1, lines 64-66). Hata discloses a dictionary of words and a word list generator coupled to the dictionary. The word list generator receives the input string and builds a word list from words stored in the dictionary corresponding to the input string. The word list generator assigns one or more prosodic environment tokens to word list entries – preferably to each entry in the word list. A reading system analyzes the word list to determine phonological features on the entries. Based on the word list, the environment token(s), and the phonological features, the reading system selects speech samples to be concatenated to supply the signal for audible synthesis. (Col. 1, line 66 – Col. 2, line 41).

Hata discloses a dictionary of digitally sampled sounds that have been recorded and stored in advance. (Col. 3, lines 42-44). The dictionary includes different samples for each possible pitch contour (i.e., prosodic environment) for each word in the dictionary. (Col. 4, lines 28-31). In addition to storing an entry for each prosodic environment of each word, the dictionary may also store all pronunciation variants of each word for each prosodic environment. (Col. 4, lines 37-55). Hata’s dictionary thus comprises multiple variations of each word entry.

Hata discloses input of text to be converted to speech. (Col. 4, lines 58-63). Hata discloses a word list generator including a prosodic environment table that identifies the possible prosodic states of the text to speech implementation. (Col. 4, lines 63-65). The word list generator builds a

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

word list comprising a word token, and a prosodic token for each word token, arranged in the order that they will be pronounced in the output speech. (Col. 5, lines 10-15). A reading system includes a phonological feature analyzer. For each word in the word list, the phonological feature analyzer evaluates the preceding and following words. Based on this information, the phonological feature analyzer selects a pre-recorded sample from the library having the corresponding prosodic and phonological characteristics. This is then added to a sample list, which is eventually output as the speech signal. (Col. 5, lines 16-31).

Regarding Claim 14, the Examiner postulates that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Sharman for specifically providing a stored speech sample corresponding to a word or other unit, as taught by Hata, for the purpose of selecting the appropriated [sic] 'granularity' or dictionary entry size to suit the specific application."

Applicant respectfully disagrees.

Again, the entire Sharman reference teaches away from processing whole words.

In order to combine the Sharman and Hata references as the Examiner has suggested – without the benefit of hindsight – one of ordinary skill in the art would have to: 1) find and read the Sharman reference; 2) understand Sharman's phoneme and diphone-based structures and operations; 3) spontaneously decide – in spite of Sharman's teaching away from processing whole words – that, out of all of Sharman's system, only its elemental and basic stored diphones needed to be improved to include larger stored speech samples in word or larger units; 4) seek out and find Hata's system; 5) disregard all of Hata's structures and methods, except for its library of pre-recorded samples for a given word and each tonal variation thereof; 6) *selectively* cull from Hata *only* its extensive library of pre-recorded samples for a given word and each tonal variation thereof; 7) disregard Sharman's

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

expressed intent to avoid excess processing; 8) augment Sharman's diphone library with Hata's extensive library of pre-recorded samples for a given word and each tonal variation thereof; 9) selectively discard Sharman's extensive teaching of, and structure and operations for, breaking words down into their constituent syllables; 10) selectively discard Sharman's structure and operations for phoneme duration assignment; 11) selectively discard Sharman's structure and operations for breath group assembly; and 12) substantially modify and supplement Sharman's syllabic, phoneme and diphone-based structures and operations to successfully process and output Hata's pre-recorded word samples.

Applicant respectfully submits that there is no teaching or motivation in either Sharman or Hata to suggest such selective and substantial modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such selective and substantial modification.

The Examiner then goes on to speculate that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Sharman in view of Hata for specifically providing a mechanism to treat a prefix or suffix as same way [sic] as a word in a dictionary, as taught by DIC, so that each stored speech sample can correspond to one of word, prefix and suffix, for the purpose of selecting the appropriated [sic] 'granularity' or dictionary entry size to suit the specific application."

Applicant respectfully disagrees.

In order to further modify the already substantially modified Sharman/Hata combination as the Examiner has suggested – without the benefit of hindsight – one of ordinary skill in the art would have to: 1) spontaneously decide that – to the extent that either Sharman or Hata disclose or suggest treatment of prefixes or suffixes – both references were somehow deficient or incomplete with regard

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

thereto; 2) seek out and find the DIC reference; 3) arbitrarily decide, based upon a dictionary entry for either a prefix or suffix, that prefixes and suffixes should be treated as words; and 4) further modify the structures and operations of the already substantially modified Sharman/Hata combination to handle prefixes and suffixes as words.

Again, Applicant respectfully submits that there is no teaching or motivation in the cited references to suggest such highly selective and highly speculative modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such modification.

The Examiner then concedes that "Sharman in view of Hata and DIC does not expressly disclose the data structure having 'a field for a frequency of a first portion of the speech sample that exceeds an amplitude threshold, and a field for a frequency of a last portion of the speech sample that exceeds an amplitude threshold'."

Nevertheless, the Examiner goes on to contend that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Sharman in view of Hata and DIC by specifically providing data structures having multiple fields for frequency or time (duration) information for processing and storing speech data, as taught by Malsheen."

Despite this highly speculative assertion, Applicant finds no suggestion or motivation in Sharman, Hata or DIC for one of ordinary skill in the art to: 1) read the Sharman reference; 2) selectively and substantially modify Sharman by Hata; 3) selectively supplement the Sharman/Hata modification with a third reference, DIC; 4) spontaneously determine, after finding no teaching or suggestion of fields for frequency of first or last portions of speech samples that exceed amplitude thresholds in a substantial and highly selective combination of three references, that "providing data structures having multiple fields for frequency or time (duration) information for processing and storing speech data" is desirable or necessary; 5) find a fourth reference, Malsheen; 6) selectively

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

cull from Malsheen only the idea of "providing data structures having multiple fields for frequency or time (duration) information for processing and storing speech data" while disregarding the remainder of Malsheen's structures and methods utilizing consonant and vowel allophones and parameters related thereto; and 7) selectively modify the Sharman/Hata/DIC combination to add structure and operations for "data structures having multiple fields for frequency or time (duration) information for processing and storing speech data".

Applicant respectfully submits that there is no teaching or motivation in either Sharman or Hata or DIC to suggest such selective and substantial modification to one of ordinary skill in the art, or to provide a reasonable expectation of successfully completing such selective and substantial modification.

Furthermore, such selective and substantial modification still fails to teach or suggest all the limitations of Claim 14.

Applicant respectfully submits that Claim 14 overcomes the rejection based upon a highly speculative and selective combination the Sharman, Hata, DIC and Malsheen references. Claim 14 is allowable. Applicant respectfully requests reconsideration and allowance of Claim 14.

o. Claims 15, 19 and 20

Claims 15, 19 and 20 depend from allowable Claim 14, and provide further limitations distinguishing over the cited references.

Applicant respectfully traverses the Examiner's interpretation of the cited references, as well as the selective hindsight combinations thereof.

Claims 15, 19 and 20 are allowable. Applicant respectfully requests reconsideration and allowance of Claims 15, 19 and 20.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

p. Claim 16

Claim 16 depends from allowable Claim 7, and provides further limitations distinguishing over the cited references.

Applicant respectfully traverses the Examiner's interpretation of the cited references, as well as the selective hindsight combination of five references.

Claim 16 is allowable. Applicant respectfully requests reconsideration and allowance of Claim 16.

q. Claim 18

Claim 18 depends from allowable Claim 12, and provides further limitations distinguishing over the cited references.

Applicant respectfully traverses the Examiner's interpretation of the cited references, as well as the selective and speculative combination of five references.

Claim 18 is allowable. Applicant respectfully requests reconsideration and allowance of Claim 18.

r. Claim 17

Claim 17 depends from allowable Claim 8, and provides further limitations distinguishing over the cited references.

Applicant respectfully traverses the Examiner's interpretation of the cited references, as well as the selective and speculative combination of six references.

Claim 17 is allowable. Applicant respectfully requests reconsideration and allowance of Claim 17.

Accordingly, the Applicant respectfully requests withdrawal of the § 103(a) rejections of Claims 1-12 and 14-23.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

IV. CONCLUSION

As a result of the foregoing, the Applicant asserts that the remaining Claims in the Application are in condition for allowance, and respectfully requests an early allowance of such Claims.

ATTORNEY DOCKET NO. 10838STUS01U (NORT10-00206)
U.S. SERIAL NO. 09/464,076
PATENT

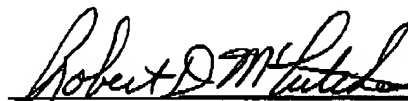
If any issues arise, or if the Examiner has any suggestions for expediting allowance of this Application, the Applicant respectfully invites the Examiner to contact the undersigned at the telephone number indicated below or at *rmccutcheon@davismunck.com*.

The Commissioner is hereby authorized to charge any additional fees connected with this communication or credit any overpayment to Davis Munck Deposit Account No. 50-0208.

Respectfully submitted,

DAVIS MUNCK, P.C.

Date: 9/8/2005



Robert D. McCutcheon
Registration No. 38,717

P.O. Drawer 800889
Dallas, Texas 75380
(972) 628-3632 (direct dial)
(972) 628-3600 (main number)
(972) 628-3616 (fax)
E-mail: *rmccutcheon@davismunck.com*